PCT/JP2004/014977

04.10.2004

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: ２００３年１０月　７日

出願番号
Application Number: 特願２００３－３４８６２１

[ST. 10/C] : ［ＪＰ２００３－３４８６２１］

出願人
Applicant(s): 松下電器産業株式会社

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

２００４年１１月　５日

特許庁長官
Commissioner, Japan Patent Office　小川　洋

出証番号　出証特２００４－３０９９９１３

【書類名】　特許願
【整理番号】　2022550317
【特記事項】　特許法第３６条の２第１項の規定による特許出願
【あて先】　特許庁長官殿
【国際特許分類】　G10L
【発明者】
　【住所又は居所】　シンガポール５３４４１５シンガポール、タイ・セン・アベニュー、ブロック１０２２、０６－３５３０番、タイ・セン・インダストリアル・エステイト、パナソニック・シンガポール研究所株式会社内
　【氏名】　コック　セン・チョン
【発明者】
　【住所又は居所】　シンガポール５３４４１５シンガポール、タイ・セン・アベニュー、ブロック１０２２、０６－３５３０番、タイ・セン・インダストリアル・エステイト、パナソニック・シンガポール研究所株式会社内
　【氏名】　スア　ホン・ネオ
【発明者】
　【住所又は居所】　大阪府門真市大字門真１００６番地　松下電器産業株式会社内
　【氏名】　田中　直也
【発明者】
　【住所又は居所】　大阪府門真市大字門真１００６番地　松下電器産業株式会社内
　【氏名】　則松　武志
【特許出願人】
　【識別番号】　000005821
　【氏名又は名称】　松下電器産業株式会社
【代理人】
　【識別番号】　100109210
　【弁理士】
　【氏名又は名称】　新居　広守
【手数料の表示】
　【予納台帳番号】　049515
　【納付金額】　35,000円
【提出物件の目録】
　【物件名】　外国語特許請求の範囲　1
　【物件名】　外国語明細書　1
　【物件名】　外国語図面　1
　【物件名】　外国語要約書　1
　【包括委任状番号】　0213583

【書類名】外国語特許請求の範囲

1 WHAT IS CLAIMED IS:

1.A method to be adopted in systems utilising time/freq grid coding of audio signals, characterised by
(a)Deriving the start time border from the end time border of the previous frame of envelope data;
(b)Detecting the most drastic transient time slot with a transient detector in the spectral data between the said start time border and the furthest allowed end time border;
(c)Finding and instantiating an actual end time border and intermediate time borders in the spectral data between the said transient time slot and the furthest allowed end time border by evaluating a signal variation criterion;
(d)Deriving the frequency resolution by evaluating the energy of every frequency band flanked by the low-resolution borders for every time segment obtained above;

2.A method according to 1, characterised in that if the number of borders allowed has been exhausted but the end time border found does not satisfy a minimum required value, expands the said intermediate borders until the minimum required value is attained;

3.A method according to 1, characterised in that more intermediate time borders can be instantiated in the spectral data between the transient time slot and the start time border by evaluating the said signal variation criterion, if the number of borders allowed has not been exhausted;

4.A method according to 1, characterised in that the said process of finding an intermediate time border first defines a temporary time segment with the previously found time border and a moving time border which moves progressively away from the said previous time border, and then evaluates the said signal variation criterion for every move the said moving time border makes.

5.A method according to 1, characterised in that the said signal variation criterion is the ratio between the minimum energy of a time slot within the said temporary time segment and the average energy of the said temporary time segment.

6.A method according to 5, characterised in that if the said computed ratio exceeds a threshold, a new intermediate or end border is instantiated according to the said moving time border to define a new time segment.

7.A method according to 2, characterised in that the said expansion of borders can occur to the time segment furthest away from the transient time slot within the said frame first, and time segments nearer to the transient time slot are considered only when the expansion of the further border has reached its syntactic limit. The said expansion of borders can also try to increase every time segment, check the signal characteristics of the new time segment formed, and applies the actual increase to the time segment that causes the least overall increase in between-border signal variations.

8.A method according to 1, characterised in that the said energy evaluation computes the ratios between the energies of the frequency bands for every time segment found. If the minimum of the ratios exceeds a threshold, a high frequency resolution is adopted; Otherwise, a low frequency resolution is adopted.

9.A method according to 8, characterised in that the said threshold is higher in a plurality of time segments immediately following the transient time border, to make it more difficult to switch to high frequency resolution in the said region.

10.A method to be adopted in bandwidth extension strategies utilising the said time/freq grid coding approach, where an analysis filterbank transforms an audio signal into a plurality of low-frequency subband signals, where portions of the said subband signals are replicated to the high-frequency region, where the said replicated subbands are divided into time segments using the said time borders information and subsequently into frequency bands using the said frequency resolutions information and subsequently modulated by the said envelope data, where a synthesis filterbank transforms the said low-frequency subband signals and the said envelope-adjusted subband signals into a bandwidth-extended, time-domain signal, characterised by

(a)Deriving the start time border from the end time border of the previous frame of envelope data;

(b)Detecting the most drastic transient time slot with a transient detector in the spectral data between the said start time border and the furthest allowed end time border;

(c)Finding and instantiating an actual end time border and intermediate time borders in the spectral data between the said transient time slot and the furthest allowed end time border by evaluating a signal variation criterion;

(d)Deriving the frequency resolution by evaluating the energy of every frequency band flanked by the low-resolution borders for every time segment obtained above;

11.A method to be adopted in systems utilising time/freq grid coding of audio signals, characterised by

(a)Deriving the start time border from the end time border of the previous frame of envelope data;

(b)Detecting the most drastic transient time slot with a transient detector in the spectral data between the said start time border and the furthest allowed end time border;

(c)Detecting which of the regions, one between the transient border and the start time border, another between the transient border and the furthest allowed end time border, has the most varying spectral data;

(d)If the said most varying spectral data is found in the region between the transient border and the furthest allowed end time border, finding and instantiating an actual end time border and intermediate time borders in the said region by evaluating a signal variation criterion;

(e)If the said most varying spectral data is found in the region between the transient border and the start time border, finding and instantiating intermediate borders in the said region by evaluating a signal variation criterion, the

n finding and instantiating an actual end time border and intermediate time borders in the other region by evaluating a signal variation criterion;

(f)Deriving the frequency resolution by evaluating the energy of every frequency band flanked by the low-resolution borders for every time segment obtained above;

12.A method to be adopted in bandwidth extension strategies utilising the said time/freq grid coding approach, where an analysis filterbank transforms an audio signal into a plurality of low-frequency subband signals, where portions of the said subband signals are replicated to the high-frequency region, where the said replicated subbands are divided into time segments using the said time borders information and subsequently into frequency bands using the said frequency resolutions information and subsequently modulated by the said envelope data, where a synthesis filterbank transforms the said low-frequency subband signals and the said envelope-adjusted subband signals into a bandwidth-extended, time-domain signal, characterised by

(a)Deriving the start time border from the end time border of the previous frame of envelope data;

(b)Detecting the most drastic transient time slot with a transient detector in the spectral data between the said start time border and the furthest allowed end time border;

(c)Detecting which of the regions, one between the transient border and the start time border, another between the transient border and the furthest allowed end time border, has the most varying spectral data;

(d)If the said most varying spectral data is found in the region between the transient border and the furthest allowed end time border, finding and instantiating an actual end time border and intermediate time borders in the said region by evaluating a signal variation criterion;

(e)If the said most varying spectral data is found in the region between the transient border and the start time border, finding and instantiating intermediate borders in the said region by evaluating a signal variation criterion, then finding and instantiating an actual end time border and intermediate time borders in the other region by evaluating a signal variation criterion;

(f)Deriving the frequency resolution by evaluating the energy of every frequency band flanked by the low-resolution borders for every time segment obtained above;

13. Software coded in programming language that provides a function achieved by the determination method according to claim 1 to 12.

14. A data recording medium for storing the software according to claim 13.

【書類名】外国語明細書

2 TITLE OF THE INVENTION

Method for Determining Time Borders and Frequency Resolutions for Spectral Envelope Coding

3 DETAILED DESCRIPTION OF THE INVENTION

3.1 Industrial Field of Utilisation

This invention introduces a systematic segmentation method to determine the time borders and frequency resolution for bandwidth extension technologies that employ a subband coding strategy, such as the Spectral Band Replication (SBR) technology.

3.2 Background and Prior Art

The objective of audio coding is to transform a digitised audio stream into a compressed representation (or bitstream) at the audio encoder, so that as high fidelity to the original source as possible is retained after the bitstream is processed at the decoder. One popular way of compression is shown in Figure 1, which shows a typical audio coding system comprising an encoder and a decoder. Module 1000 divides the audio signal in time domain into consecutive frames, module 1010 transforms each frame of audio signal into frequency domain and module 1020 quantizes the spectrum up to a certain frequency (known as the bandwidth) at the encoder. One possible way for module 1010 to transform the audio signal into frequency domain is the time/frequency grid approach as shown in Figure 18, where a filterbank is employed to split an audio signal into multiple subbands, each representing a portion of the signal within a narrow frequency range in time domain. At the decoder, the audio spectrum is de-quantized by module 1030 and inverse-transformed by module 1040 back into audio frames. The audio frames are then appropriately assembled by module 1050 to form a continuous audio stream.

As the bitrate (number of bits per second) of- coding decreases, more sacrifice has to be made to the bandwidth by not coding the high-frequency portion, as it is deemed not as perceptually important as the low frequency portion. The consequence is that some high-frequency tones, and harmonics of the low-frequency tones are shut down. Figure 2 illustrates the above band-limiting operation, where 2020 indicates the resultant bandwidth of the coded audio.

The objective of bandwidth extension is to recover the high-frequency portion, by coding them using very few additional bits. One example of such a technique is the Spectral Band Replication (SBR) method (International Patent Publication WO98/57436), which is now an MPEG standard (ISO/IEC 14496-3, 2001 AMD1). Figure 3 illustrates one possible encoder structure for SBR that is relevant to this invention. At the outset, the audio signal is band-splitted into N subbands with N subband filters at the 'analysis filterbank' 3010, each capturing a part of the signal's frequency spectrum. The N signals produced by the filters are decimated to remove redundancy. The bandwidth extension coder 3020 extracts some

information from the filter outputs so that at the decoder, the low-frequency subbands can use the information to extend the bandwidth of the audio signal. The bandwidth extension information is then multiplexed at 3030 with the output of the core audio encoder 3000 to form a bitstream. A nominal SBR frame consists of L outputs from each subband filter.

Figure 4 illustrates the decoder for the SBR method that is relevant to this invention. At the outset, a bitstream is demultiplexed at 4000 to become the core audio bitstream and the bandwidth extension bitstream. The core decoder 4010 decodes the core audio bitstream to produce the band-limited audio signal in time domain. The band-limited audio signal is then band-splitted into M subbands with M subband filters of the 'analysis filterbank' 4020. Higher-frequency subbands are synthesized using the bandwidth extension information at this subband level. The new higher-frequency subbands, as well as the lower-frequency subbands, are up-sampled and assembled with an N-filter 'synthesis filterbank' 4040 to output the final bandwidth-extended signal.

The output from the analysis filterbank 3010 can be viewed as a time/frequency grid representation of the audio signal as shown in Figure 18. As part of the bandwith extension information, the time frequency representation is to be divided first in the time direction into 'time segments' and then in the frequency direction into 'frequency bands'. For each frequency band, its average energy is computed, quantized and coded. This process is known as spectral envelope coding. Figure 5 illustrates such a segmentation process, and is fully described in International Patent Publication WO01/26095A1. In the figure, 5010 depicts segmentation in the time direction, and 5020 depicts segmentation in the frequency direction At the decoder, the data generated by this process is used to shape the energy of the synthesised high-frequency bands, so that it takes on the same energy envelope as the original audio signal. Without proper segmentation, low-energy areas would be forced to share the same average energy value as the large-energy areas. This would in turn lead to erroneous amplification at the decoder, which is a common source of audible artefacts.

Each SBR frame is partitioned in the time direction into time segments using 'borders'. The prior art describes the method of using 'fixed' and 'variable' borders to achieve effective spectral envelope coding. Refer to Figure 6, the fixed borders 6060, 6070 and 6100 coincide with the borders 6010, 6020 and 6050 of the nominal SBR frames, whereas the variable borders 6080 and 6090 of the current frame is allowed to encroach into the next nominal SBR frame. The start border and the end border of a 'variable SBR frame' can either be a fixed border or a variable border. If the start border and end border are both fixed borders, the variable SBR frame coincides with the nominal SBR frame. The end border of the current SBR frame automatically becomes the start border of the next SBR frame.

Between the start border and end border, the SBR frame is further partitioned into several time segments by intermediate borders according to the prior art. If the start border and end border are both fixed borders, the SBR frame is partitioned into uniform time segments. This is known as the FIXFIX frame in the prior art (i.e a FIX border as the start border and a FIX border as the end border), and is depicted in Figure 7, where 7010 is the start border and 7020 is the end border. If a threshold detector finds a transient region in the current SBR frame, its end border will become a 'variable' border that must be equal to or greater than the next nominal SBR frame. This is the so-called FIXVAR frame shown in Figure 8. It has a FIX border as the start border 8010 and a VAR border as the end border 8050. The intermediate borders 8020, 8030 and 8040 are specified relative to one another or the variable border, where d_0, d_1, d_2, etc are the relative border distances. According to Figure 8, the first relative

distance d_0 must start with the variable border. Subsequent relative distances start with the previously determined intermediate borders.

Since the end border of the current SBR frame automatically becomes the start border of the next SBR frame, it's possible for an SBR frame to have two variable borders in case of transient behaviours in successive SBR frames. This is known as the VARVAR frame as illustrated in Figure 9. For a VARVAR frame, the intermediate borders can be specified as relative to either one of the variable borders. In the figure, intermediate border 9020 is relative to start border 9010, whereas intermediate borders 9030, 9040, 9050 are relative to each other or the variable end border 9060.

Finally, if the transient detector cannot find any transient in the current SBR frame, but it begins with a variable border, it will still adopt a fixed border as its end border. This is the final frame class introduced in the prior art and is known as a VARFIX frame, as illustrated in Figure 10, where 10010 is a variable start border and 10050 is a fixed end border. 10020, 10030 and 10040 constitute the intermediate borders progressively derived from d_0, d_1 and d_2.

To reduce bit consumption, the relative border distances between an intermediate border and a variable border can only take on a few pre-determined sizes.

After marking a plurality of time segments with the above-mentioned borders, each time segment, flanked by two borders, is to be divided in the frequency direction into frequency bands. The exact spectral borders are derived using criteria that are irrelevant to this invention. Two possible resolutions can be specified: high or low. Figure 11 shows the border relationship between a high-resolution division and a low-resolution division. The borders of the low-resolution divisions are the alternate borders of the high-resolution division.

3.3 Problems

For current SBR frame, upon the determination of the start border based on the end border of the previous SBR frame, and the determination of a transient border using a threshold detector, a method is needed to determine the end border, and all intermediate borders.

The problem is not straightforward because, as mentioned, all intermediate borders d_i are to be specified relative to one another or the variable borders, and all relative distances can only take on a few pre-determined sizes, $d_i \in \{D_1, D_2, D_3, D_4\}$, with $0<D_1<D_2<D_3<D_4$. Moreover, only a syntactically pre-determined number of intermediate borders are permitted. For the FIXVAR and VARVAR frame type, the end border must be equal to or greater than the nominal SBR border. A systematic method is needed to encompass all constraints imposed.

The default spectral coding strategy adopted by the prior art resorts to low temporal resolution but high spectral resolution (i.e. few time segments but more frequency bands). When a transient is detected, the prior art switches to high temporal resolution but low spectral resolution (i.e. more time segments but less frequency bands) to code the region after the transient. The objective for switching the degrees of resolution is to account for the fact that a transient tends to exhibit more temporal variation than spectral variation. Lowering the frequency resolution can help curb a sudden surge in bit consumption. However, this method is not sufficient if the post-transient region exhibits a high degree of spectral variation that warrants a higher resolution, such as the case of a sudden burst of a tonal signal.

3.4 Solution to the Problems

3.4.1 Determination of Time Borders

To determine the time borders, this invention discloses a systematic method to determine the end border and all intermediate borders while taking into account all syntactic constraints imposed by the decoder.

Like in the prior art, the frame type for the current SBR frame is determined according to the type of end border of the previous frame, as well as the presence of a transient in the current SBR frame. The start border is also determined according to the end border of the previous SBR frame.

For a FIXFIX frame, a low time resolution setting is used.

For a FIXVAR frame and a VARVAR frame, a search for possible intermediate borders is first conducted in the region after the transient time slot. The end border is also determined at this stage. Then, another search is conducted in the region before the transient time slot for possible intermediate borders, if the first stage hasn't already exhausted the maximum number of borders allowed.

For the VARFIX frame, only one search needs to be conducted, in the whole region flanked by a variable start border and a fixed end border.

All of the above are accomplished with two Forward Search operations and one Backward Search operation. They employ the same principle, which is based on evaluating the signal variation of a time segment, but with minor variations to suit the scenarios in which they are applied.

3.4.2 Determination of Frequency Resolution

To determining the frequency resolution, this invention discloses an adaptive method that objectively assesses the energy variation in the spectral direction.

Since the borders of low-resolution division are the alternate borders of high-resolution division, a high resolution is first assumed and average energies are computed for each frequency band. For every pair of frequency bands flanked by the low-resolution borders, the ratio of energies is computed. If the minimum of all energy differences computed for the entire time segment exceeds a pre-determined threshold, a high-frequency resolution is adopted. Otherwise, a low-frequency resolution is adopted. Noting the importance of giving employing high temporal resolution in the post-transient region, the method applies a stricter criterion for the adoption of high frequency resolution in this region.

3.5 Embodiments

The below methods are examples explained in the context of SBR. However, their applicability extends to any embodiments utilizing spectral envelope coding based on time/frequency grid.

3.5.1 Determination of Time Borders

The embodiment for the determination of time borders is presented as a series of flowcharts shown in Figure 12-15.

3.5.1.1 Overview

Figure 12 shows an overview of the overall time border determination operation. 12010 sets the first border *border*[0] to the end border of the previous SBR frame. It also initialises the border counter *noBorder* to 1. 12020 activates the transient detector for the current frame, to check for the most drastic transient behaviour from border[0] to (next nominal SBR border + *V*), where *V* is the amount of transgression into the next SBR frame allowed by the syntax.

If a transient is found, 12030 checks the end border of the previous SBR frame for its type. If it's a FIX border, the current frame becomes a FIXVAR type in 12050; If it's a VAR border, the current frame becomes a VARVAR type in 12090. In either case, the transient border is registered in border[1] and *noBorder* is incremented.

If a transient is not found, 12040 checks the end border of the previous SBR frame for its type. If it's a FIX border, the current frame becomes a FIXFIX type in 12130; If it's a VAR border, the current frame becomes a VARFIX type in 12150.

If the current frame is FIXVAR, 12060 checks the region between the said transient and (next nominal SBR border + *V*) for possible need for intermediate borders. The Forward Search (Type I) method to be described in 3.5.1.2 is used for this purpose. At the end of Forward Search, *noBorder* is checked in 12070. If *noBorder* is found to be below the maximum allowed number of borders *MaxBorder*, 12080 uses a Backward Search method to check the region between the said transient and the start border and instantiate more intermediate borders if necessary. The above sequence of operations prioritises the post-transient region in finding intermediate borders.

If the current frame is VARVAR, 12100 checks the region between the said transient and (next nominal SBR border + *V*) for possible need for intermediate borders using the same Forward Search (Type I) method to be described in 3.5.1.2. At the end of Forward Search, *noBorder* is checked in 12110. If *noBorder* is found to be below the maximum allowed number of borders *MaxBorder*, 12120 uses another Forward Search method (Type II) to check the region between the said transient and the start border and instantiate more intermediate borders if necessary. Again, the above sequence of operations prioritises the post-transient region in finding intermediate borders.

If the current frame is FIXFIX, 12140 opts for a low temporal resolution setting. More is discussed in 3.5.2.

If the current frame is VARFIX, 12160 checks the region between the start border and the next nominal SBR frame border for possible need for intermediate borders. The aforementioned Forward Search (Type I) method is used for this purpose.

The four branches of operations culminate in 12170 which sorts the borders generated in ascending order for later processing.

Figure 17 depicts the employment of the three search types in the four frame types, where 17010 and 17020 denote the forward search (type I) operation, 17040 and 17050 denote the forward search (type II) operation, and 17030 denotes the backward search operation.

The post-transient region is prioritised in the intermediate border determination process in the above embodiment, however, it is also possible to select which of the regions should be prioritised by evaluating signal variations. If the signal variation is lager in the pre-transient region, the pre-transient region is prioritised, and if it is larger in the post-transient region, vice versa.

3.5.1.2 Forward Search (Type I)

This Forward Search (Type I) method is designed for a region that starts with a transient and ends with a variable border which is yet to be determined. Its objective is to determine the intermediate borders and the end border. Three input parameters, *border1*, *border2* and *noBorderLimit* must be initialised according to 12060 and 12100 of Figure 12 to delineate the search zone (between *border1* and *border2*), and the maximum number of borders permitted.

The flowchart of this method is shown in Figure 13. The method uses two intermediate variables i and j to track the left and the right border of a time segment. k is used to index the relative border distance D_k for the current time segment. 13010 initialises i to *border1* and k to 2. 13020 checks whether i is still below the nominal SBR frame border and the *noBorder* hasn't exceeded the *noBorderLimit*. If the condition is passed, more intermediate borders can still be instantiated, so 13030 sets the next possible edge of the current time segment, j, to $i+D_2$. 13040 checks j on whether its value exceeds *border2*. If it does, then D_k is not a usable relative border distance. The method reverts to the previous relative border distance, D_{k-1} by subtracting 1 from k in 13090 and registering a new border at $i+D_k$. The number of borders is updated by incrementing *noBorder*. If the method arrives at 13100 via the 'no' decision path of 13040, then the border just registered would later become the variable end border of this SBR frame.

On the other hand, if 13040 produces a 'yes' decision, it proceeds to evaluate a signal variation criterion to find out whether a new border is necessary. However, if D_k is already the maximum allowed relative border distance (D_4 in this example), as reflected in 13050, then the signal variation criterion needs not be evaluated as a new border becomes compulsory. It would branch directly to 13100 to register the new border.

If D_k is not D_4 yet, then a variable *peak_ratio* is evaluated in 13060 for the region between i and j-1. One possible criterion for a new intermediate border can be based on checking the ratio of the energy of each time slot to the average energy of the entire time segment. It's carried out in 13070 as shown:

$$peak\_ratio = \min\left\{\frac{ET_m}{\overline{ET}}\right\} > Tr_1 \qquad \text{, for } i \le m \le j\text{-}1$$

where,

ET_m is the energy of time slot m,
$\overline{ET}$ is the average energy of all time slots, computed from i to j-1
Tr_1 is a pre-determined threshold value.

Another possible signal variation criterion can be based on comparing the largest and smallest energy as follows:

$$peak\_ratio = \frac{\text{largest } ET_m \text{ of all time slots from } i \text{ to } j\text{-}1}{\text{smallest } ET_m \text{ of all time slots from } i \text{ to } j\text{-}1} > Tr_1$$

Lastly, the signal variation criterion can be based on comparing the largest and smallest absolute amplitudes as follows:

$$peak\_ratio = \frac{\text{largest absolute amplitude of all time slots from } i \text{ to } j\text{-}1}{\text{smallest absolute amplitude of all time slots from } i \text{ to } j\text{-}1} > Tr_1$$

If *peak_ratio* exceeds a threshold Tr_1, then the large signal variation warrants a new border. However, as the current D_k causes the large signal variation, D_{k-1} should be the desired relative border distance. As a result, the value of k is decremented in 13090 and a new border is registered in 13100.

If *peak_ratio* is not above the threshold Tr_1, the signal variation is considered fairly even, so a larger D_k is attempted by first incrementing k followed by adjusting j in 13080.

The process repeats until finally 13020 returns a 'no' decision. It then proceeds to 13110 to check whether despite using up all the *noBorderLimit*, the last border (which would become the variable end border) is still below the nominal SBR frame border. This is an important consideration because the SBR syntax requires that the end border be equal or greater than the nominal SBR frame border. If not the case, the operation safely terminates. If it's the case, the method begins a process of expanding the relative border distances until the last border satisfies the above requirement.

One possible method to expand the relative border distances is by sacrificing the relative border distance that's the furthest away from the transient border first. Starting from 13120, i is initialised to index the last border. 13130 checks the relative border distance between *border*[i] and *border*[i-1]. If the difference is not less than D_4, this relative border distance cannot be expanded, so i is decremented so that the relative border distance between *border*[i-1] and *border*[i-2] is checked subsequently. However, if the difference is below D_4, the relative distance between *border*[i] and *border*[i-1] is expanded in 13160. The process is repeated until the last border is greater or equal to the nominal SBR frame border as verified in 13170.

Another method of expanding the relative border distances is more computationally intensive. It tries to increase every relative border distance between borders, check the signal characteristics between the new borders, and applies the actual increase to the relative border distance that causes the least overall increase in between-border signal variations. Then the operation is repeated until the end border becomes equal or greater than the nominal SBR frame border. However, from experience, the region that is least varying is also the region that is furthest away from the transient border, because if the region near the transient border were the most varying, this characteristic would have already been captured by the presence of closely spaced intermediate borders near the transient border.

3.5.1.3 Forward Search (Type II)

This Forward Search (Type II) method is designed for a region that starts with a variable or fixed border, and ends with a border that has already been determined, such as the transient border or a fixed border. Unlike the Type I method, its objective is to determine the intermediate borders only. Three input parameters, *border1*, *border2* and *noBorderLimit* must be initialised according to 12120 and 12160 of Figure 12 to delineate the search zone and the maximum number of borders permitted.

The flowchart of this method is shown in Figure 14. In principle, the two search methods are the same. Therefore, operations 14010 to 14100 are almost identical to operations 13010 to 13100 of Figure 13, with a few exceptions:

In 14020, instead of checking whether the leading edge of the current time segment is below the next nominal SBR frame, the new constraint is for the leading edge to be below $border2-D_2$.

If 14020 returns a 'no' decision, the operation terminates. There is no need for the operation to expand some relative border distances (i.e. unlike 13110 onwards in Figure 13) because an end border needs not be found.

Similarly, in 14040, if the trailing edge of the current time segment exceeds *border2*, it terminates right away as opposed to registering a new border at $i+D_{k-1}$ (i.e. the branching from 13040 to 13090 in Figure 13) as an end border is not necessary.

In 14100, the *peak_ratio* of a new border has to be stored when it is instantiated. This is to facilitate 14110, which removes redundant borders. Redundant borders are sometimes created because the size allowed for the current time segment has reached a maximum. Since the border locations are to be specified relative to each other, this border is necessary if more borders are to be created subsequently. However, if this is the last border, it can be removed without causing any problem.

3.5.1.4 Backward Search

This backward Search method is designed for a region that starts with a transient and ends with a start border. Three input parameters, *border1*, *border2* and *noBorderLimit* must be initialised according to 12080 of Figure 12 to delineate the search zone and the maximum number of borders permitted.

The flowchart of this method is shown in Figure 15. In principle, the method is the same as Forward Search (Type II). Therefore, operations 15010 to 15110 are almost identical to operation 14010 to 14110, except that the operations are performed in the reverse direction: Instead of incrementing j relative to i, backward searching decrements j relative to i. Specifically,

Instead of $i \leq border2-D_2$ in 14020, there is $i \geq border2+D_2$ in 15020 because i will get increasingly closer to the start border (i.e. *border2*).

Instead of $j \leq border2$ in 14040, there is $j \geq border2$ in 15040 for the same reason mentioned above.

Instead of computing *peak_ratio* for time slots i to j-1 in 14060, 15060 computes *peak_ratio* for time slots j to i-1.

Instead of computing $j=i+D_k$ in 14030 and 14080 , there is $j=i-D_k$ in 15030 and 15080.

Finally, instead of computing $i=i+D_k$ in 14100, there is $i=i-D_k$ in 15100.

3.5.2 Low Temporal Resolution for FIXFIX

A FIXFIX frame has no transient characteristics in its vicinity, so it's logical to use very few time borders to save coding bits. For SBR, the time/frequency representation for a FIXFIX frame is uniformly divided based on the number of borders chosen. A simple method to choose the number of borders is to try out the lowest number of borders and evaluate the *peak_ratio* of the time segments formed. If any of the *peak_ratio*'s exceeds a certain threshold, a larger number of borders is tried, and the evaluation of *peak_ratio* for each time segment formed is repeated. The process terminates when the *peak_ratio*'s of all time segments formed are below a threshold, or when the maximum number of borders has been reached.

3.5.3 Determination of Frequency Resolution

The embodiment for the determination of frequency resolution is illustrated by way of an example shown in Figure 16. The borders of low-resolution division are the alternate borders of high-resolution division.

Initially, the average energy for every frequency band in a time segment is computed, assuming that a high frequency resolution is adopted. The average energy is denoted by E_i.

If the high frequency resolution is even, then satisfying the following condition will lead to the selection of high frequency resolution; Otherwise, the low frequency resolution will be selected:

$$\min\left\{\frac{E_{2i+1}}{E_{2i}}\right\} > Tr_2 \qquad \text{, for } i=0,1,2 \ldots$$

If the high frequency resolution is odd, then satisfying the following condition will lead to the selection of high frequency resolution; Otherwise, the low frequency resolution will be selected:

$$\min\left\{\frac{E_{2i+2}}{E_{2i+1}}\right\} > Tr_2 \qquad \text{, for } i=0,1,2 \ldots$$

where

$$Tr_2 = \begin{cases} FREQ\_RES\_THRESHOLD_1, & \text{for the first } n \text{ time segments after a threshold border} \\ FREQ\_RES\_THRESHOLD_2, & \text{otherwise} \end{cases}$$

and $FREQ\_RES\_THRESHOLD_2 > FREQ\_RES\_THRESHOLD_1$. This implies that for the *n* time segments after a threshold time slot, it's harder to adopt a high frequency resolution because a higher time resolution is favoured.

While the average energy is used for the determination in the above embodiment, any other parameter like amplitude information, which represents signal variation, can be used instead.

3.6 Effects of the Invention

The time border determination method successfully provides a systematic method to perform frame segmentation in the pre- and post transient regions by evaluating the change of energy in time. It provides good sound quality by emphasizing the post-transient region over the pre-transient region, and the region closest to the onset of transient over the region further away, while taking into considerations all syntactic constraints imposed. The adaptive frequency resolution determination method helps to check the energy distribution in the frequency direction in the post-transient region. It resorts to high-resolution segmentation if a large variation in the energy distribution is detected. Together, the two methods of the invention realise a good and easily implemented strategy for segmentation of the time-frequency representation of SBR technology.

【書類名】外国語図面
【図１】

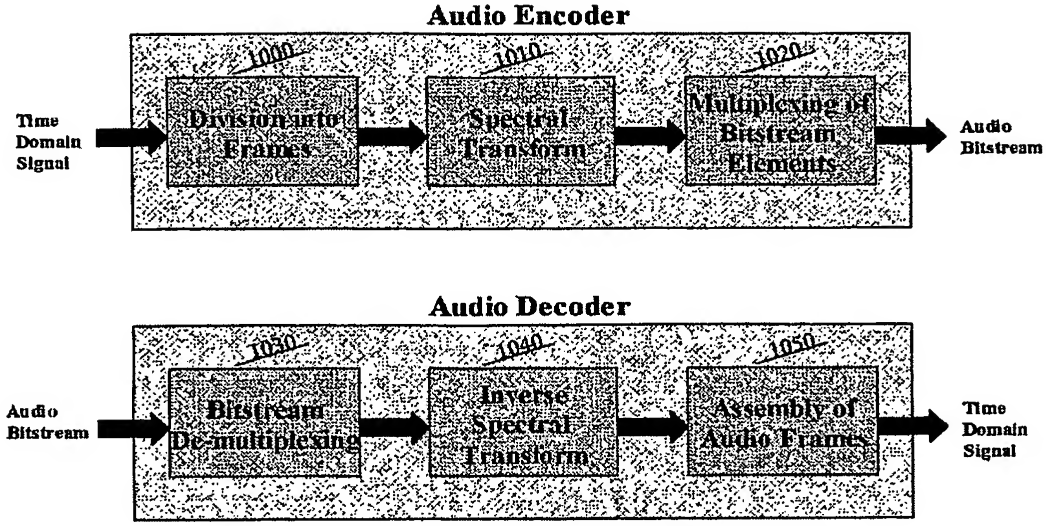

Figure 1: A typical audio coding system

【図２】

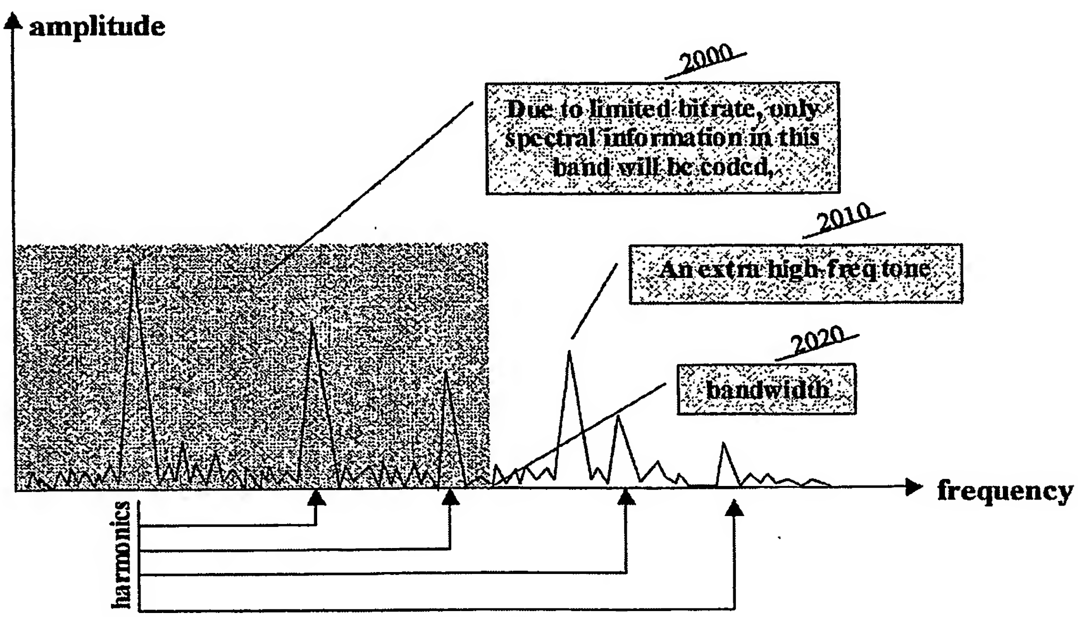

Figure 2: Limitation of bandwidth owing to bitrate consideration causes a loss of some high-frequency tones and harmonics

【図３】

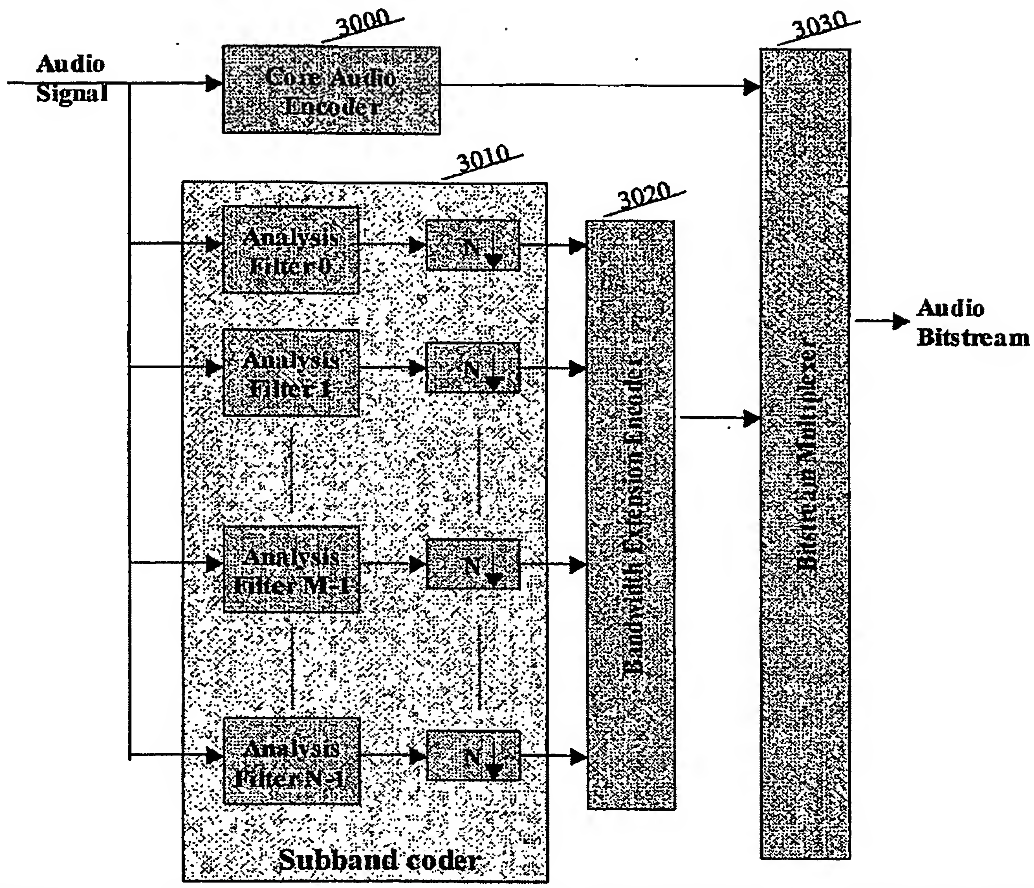

Figure 3:　A possible encoder of a subband coding scheme for bandwidth extension

【図４】

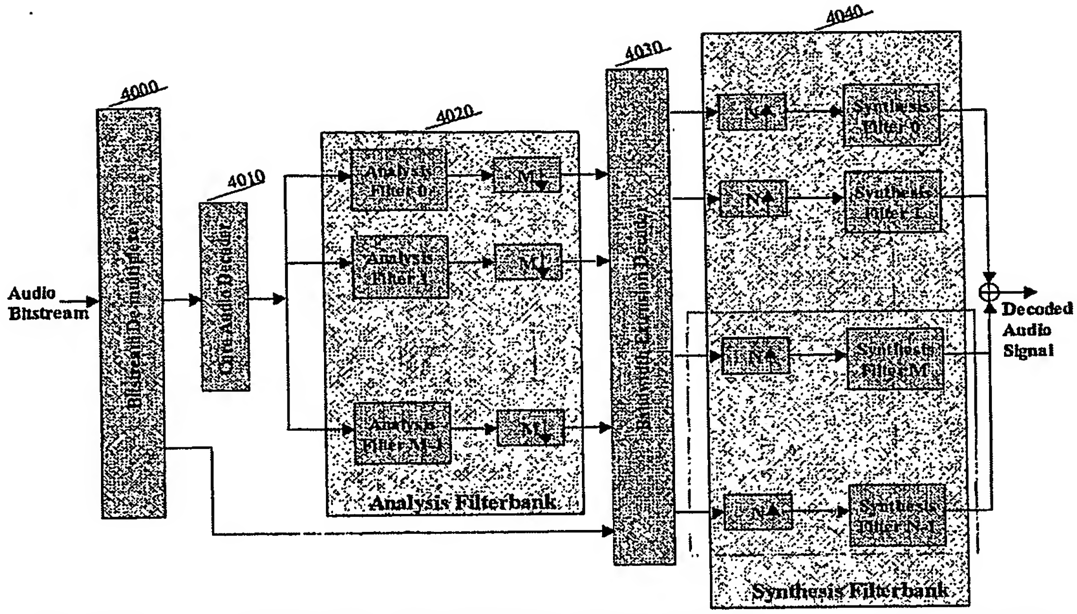

Figure 4:　　The decoder of a subband coding scheme for bandwidth extension

【図５】

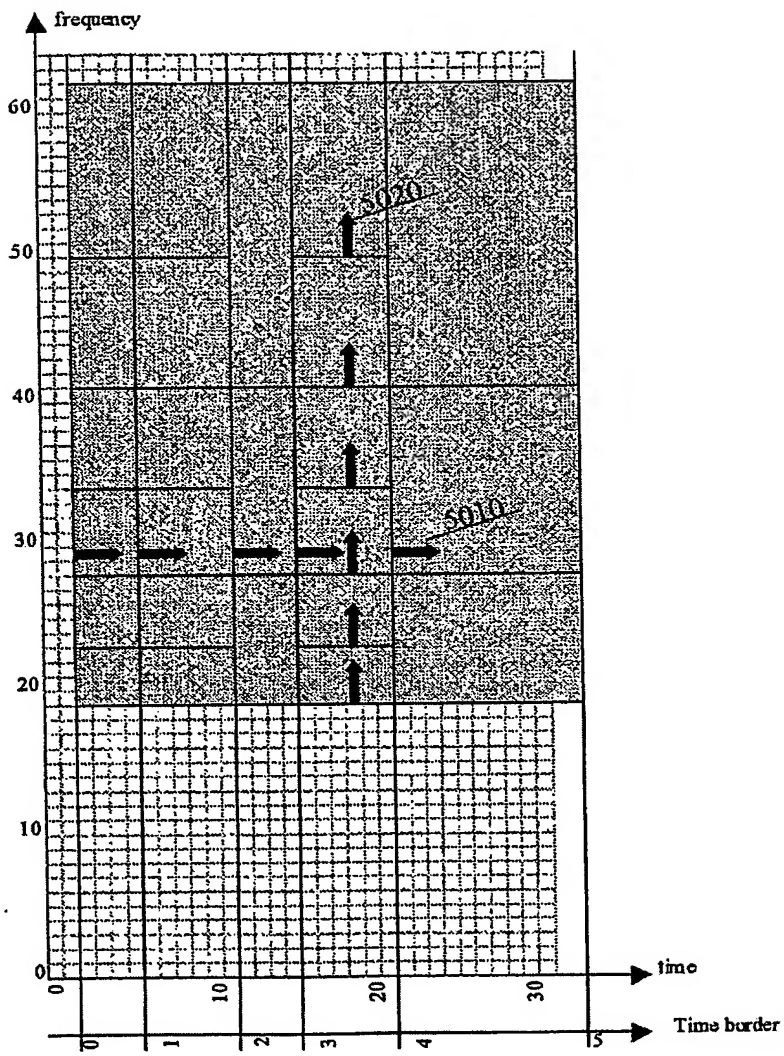

Figure 5: Segmentation in the time and frequency direction.

【図６】

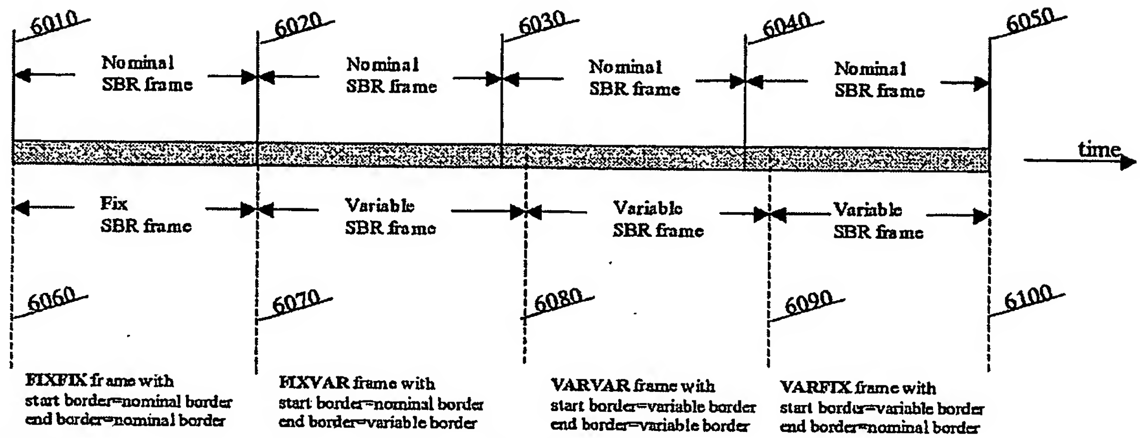

Figure 6: Border relationships between four frame types: FIXFIX, FIXVAR, VARFIX, VARVAR.

【図７】

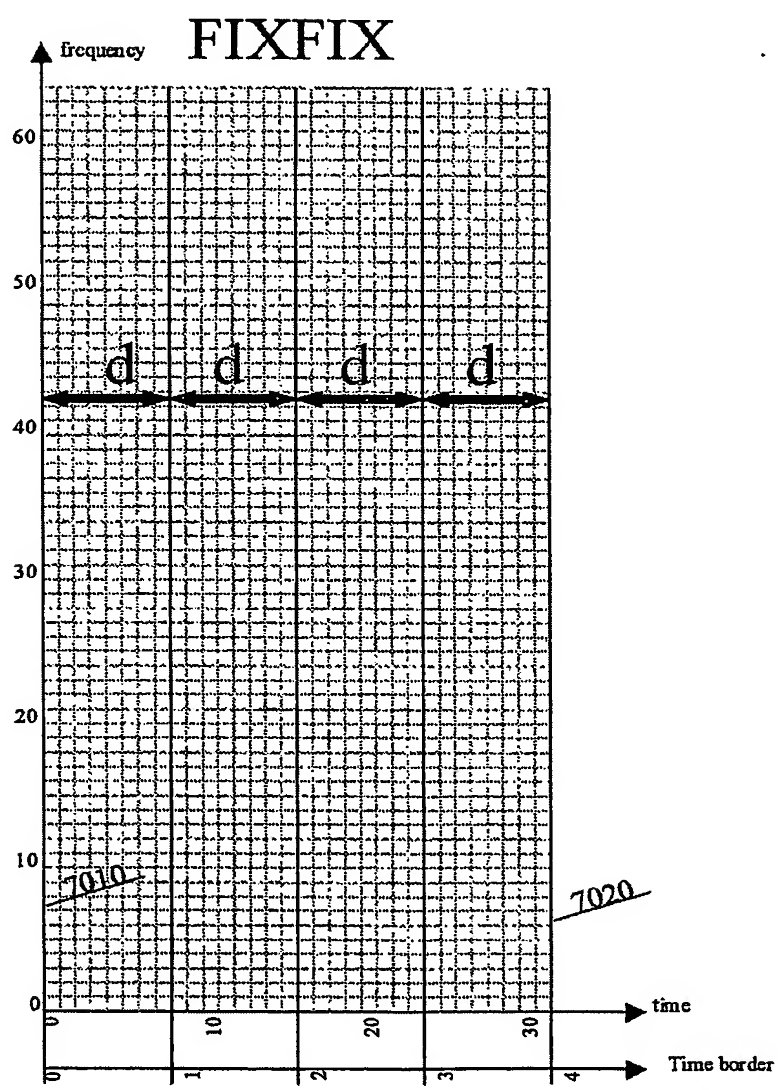

Figure 7:　A FIXFIX frame with fixed start and end borders.

【図８】

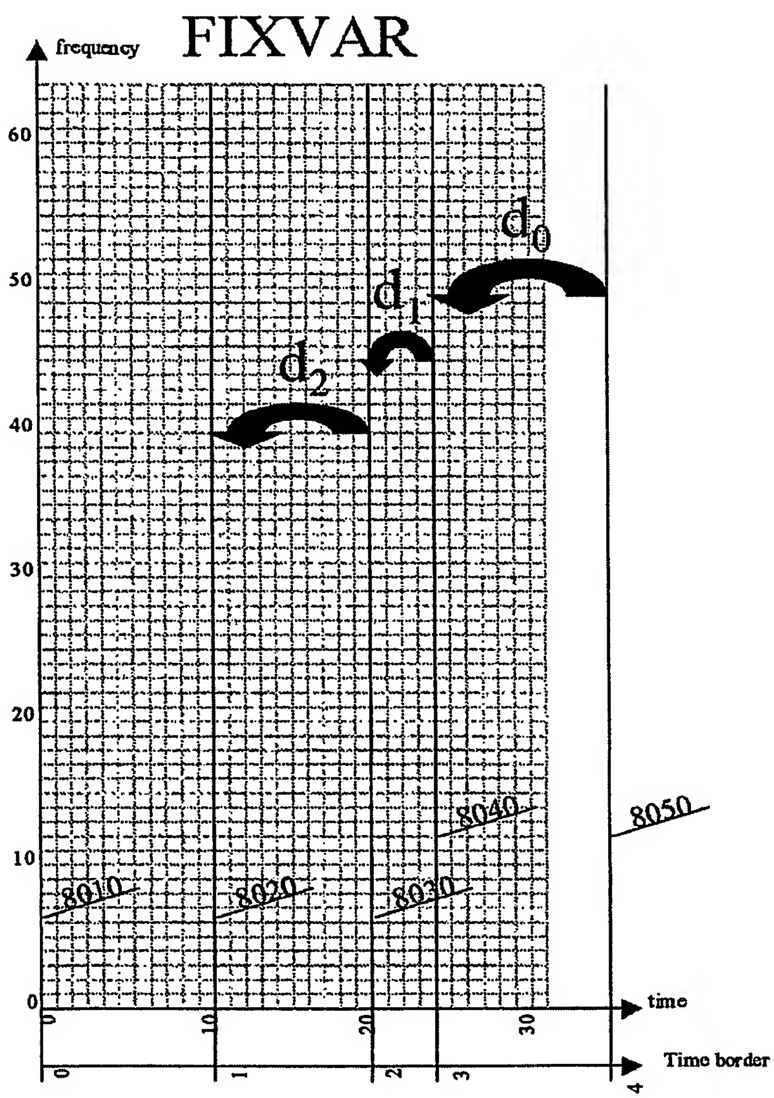

Figure 8: A FIXVAR frame with a fixed start border, a variable end border greater than the nominal SBR frame border, and some intermediate borders specified relative to the end border or each other.

【図９】

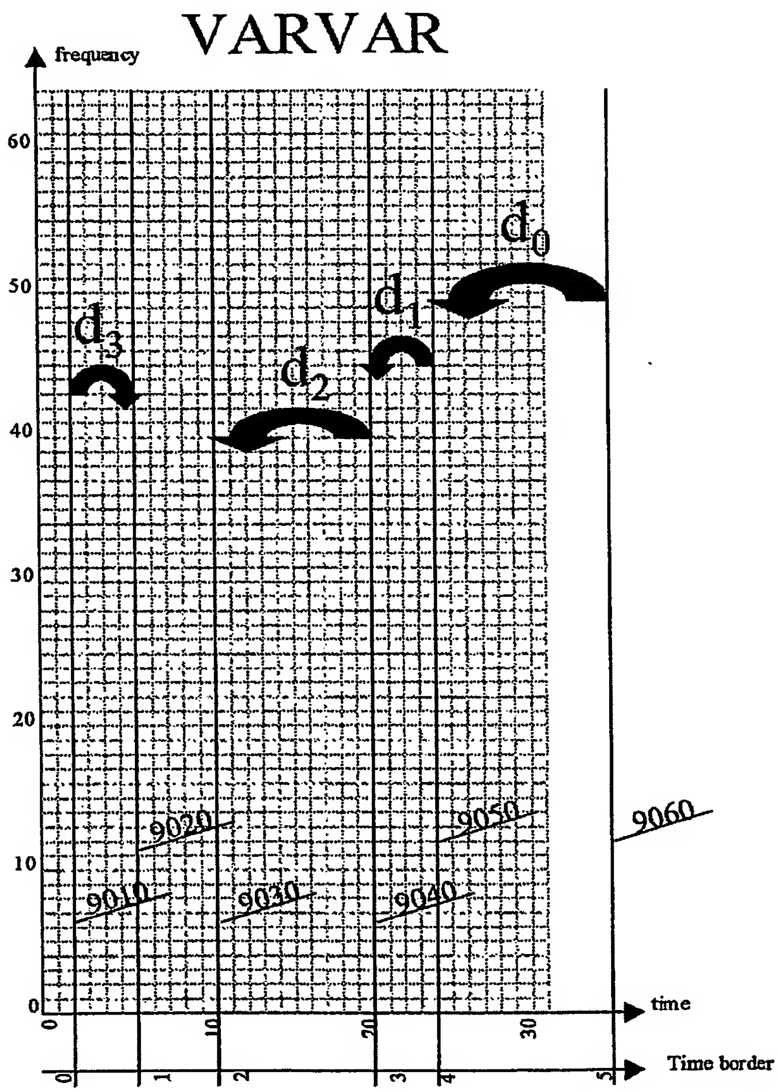

Figure 9: A VARVAR frame with a variable start border, a variable end border greater than the nominal SBR frame border, and some intermediate borders specified relative to the two end borders or each other.

【図１０】

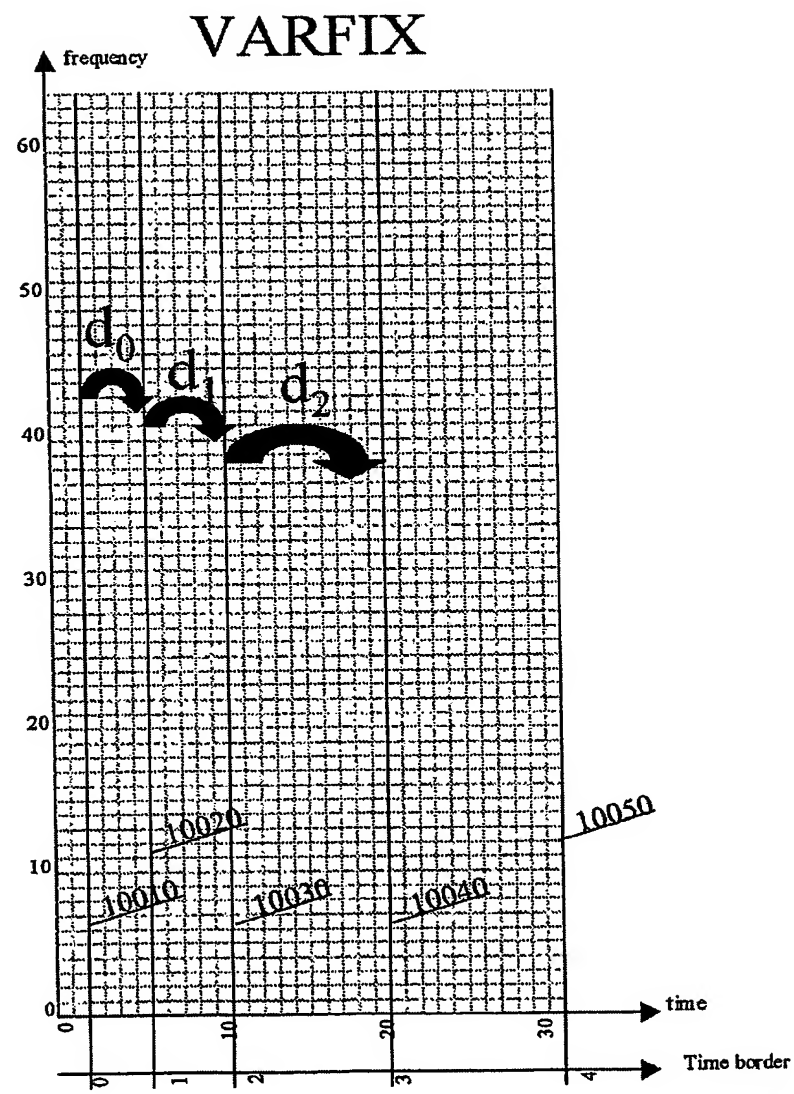

Figure 10: A VARFIX frame with a variable start border, a fixed end border, and some intermediate borders specified relative to the start border or each other.

【図１１】

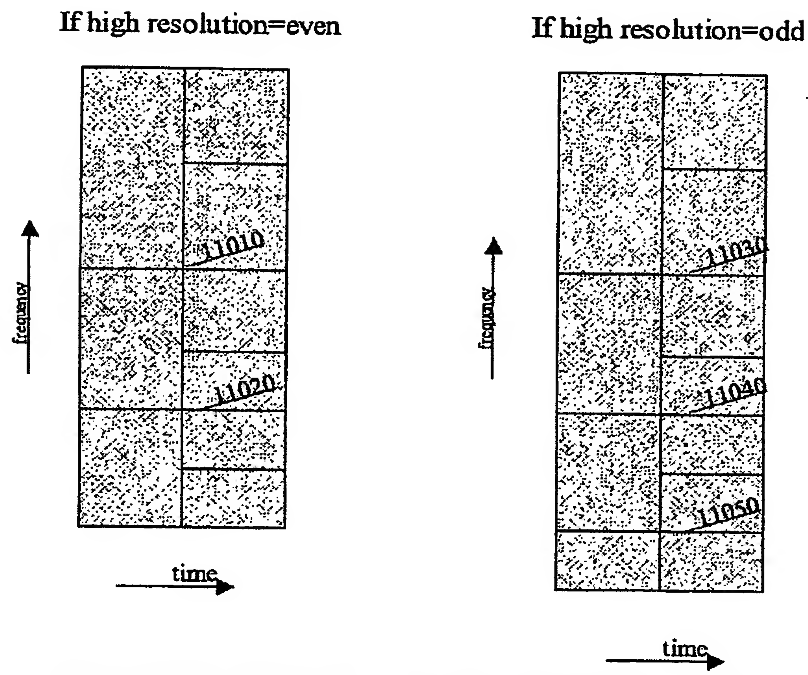

Figure 11: Border relationship between a high-resolution time segment and a low-resolution time segment

【図１２】

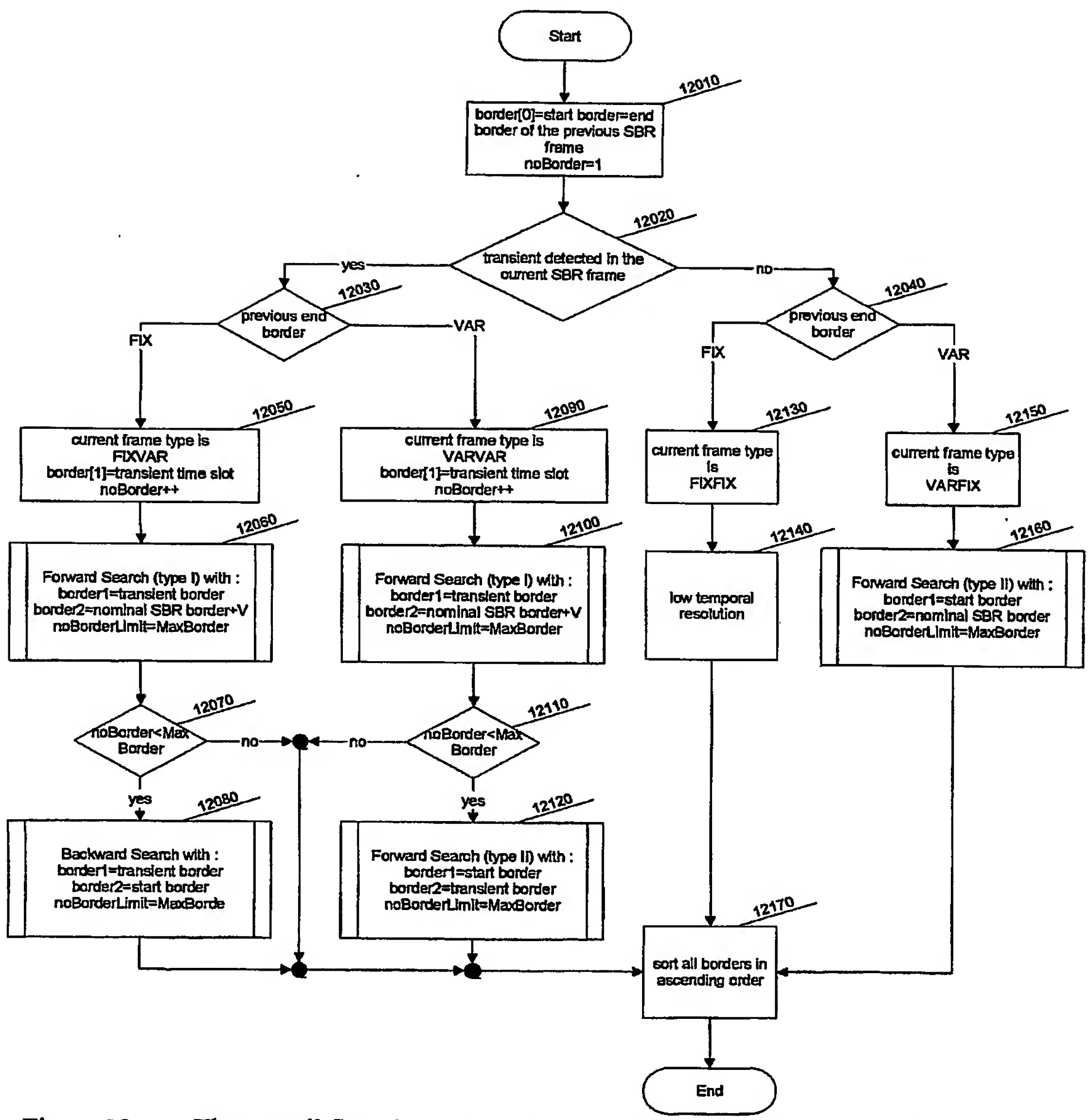

Figure 12: The overall flowchart of the time border determination part of the invention

【図１３】

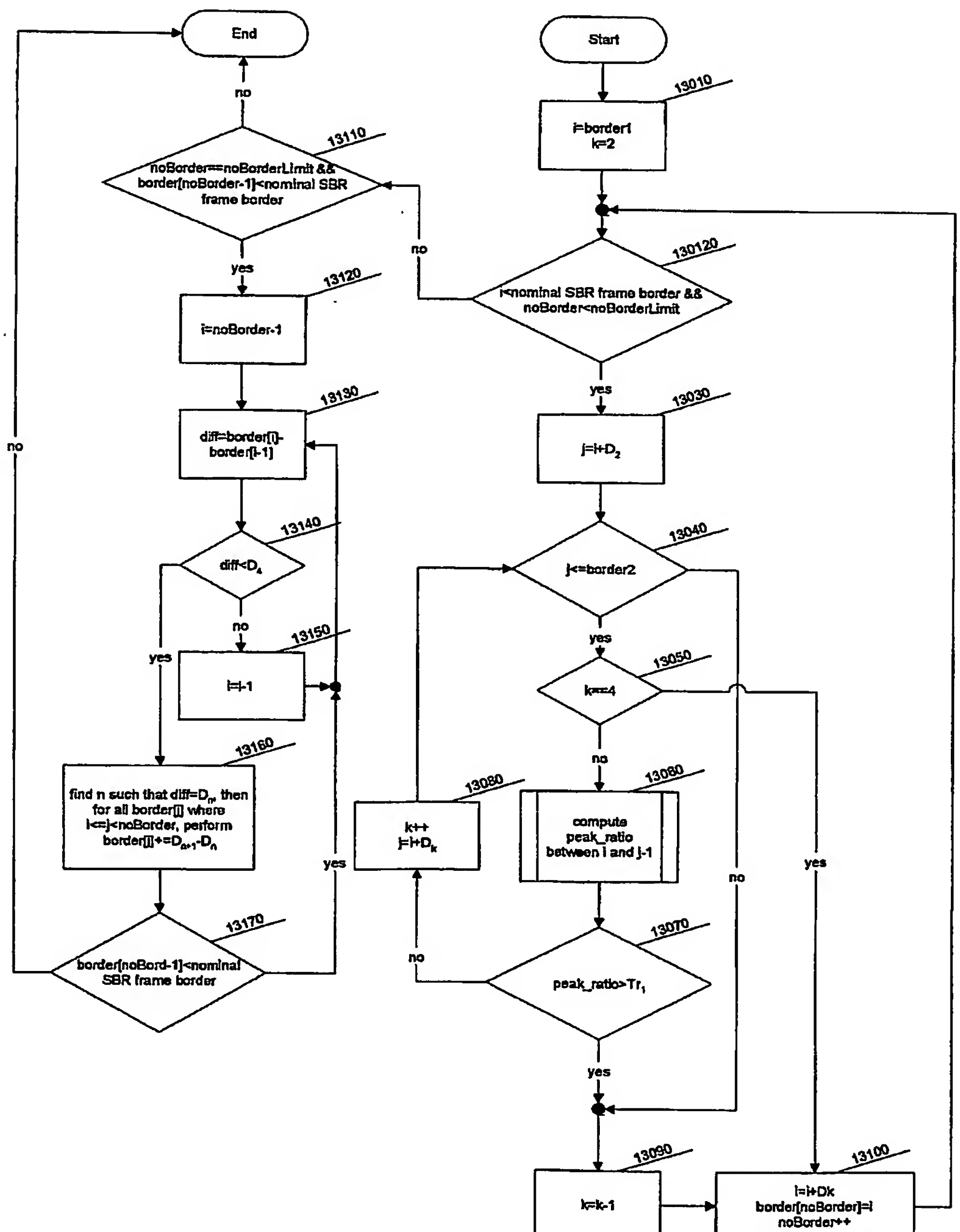

Figure 13:　The flowchart of the Forward Search (Type I) operation

【図１４】

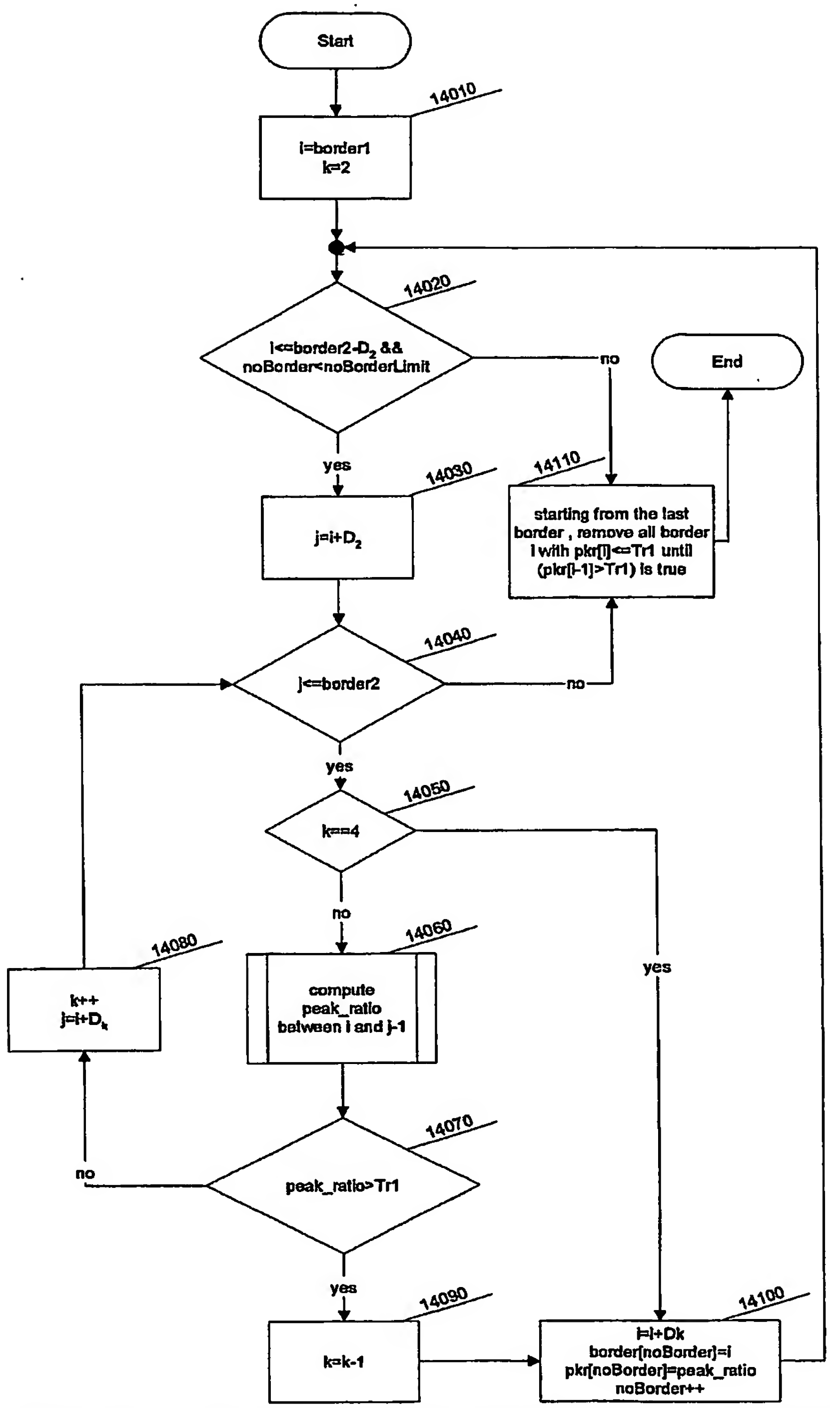

Figure 14:　The flowchart of the Forward Search (Type II) operation

【図１５】

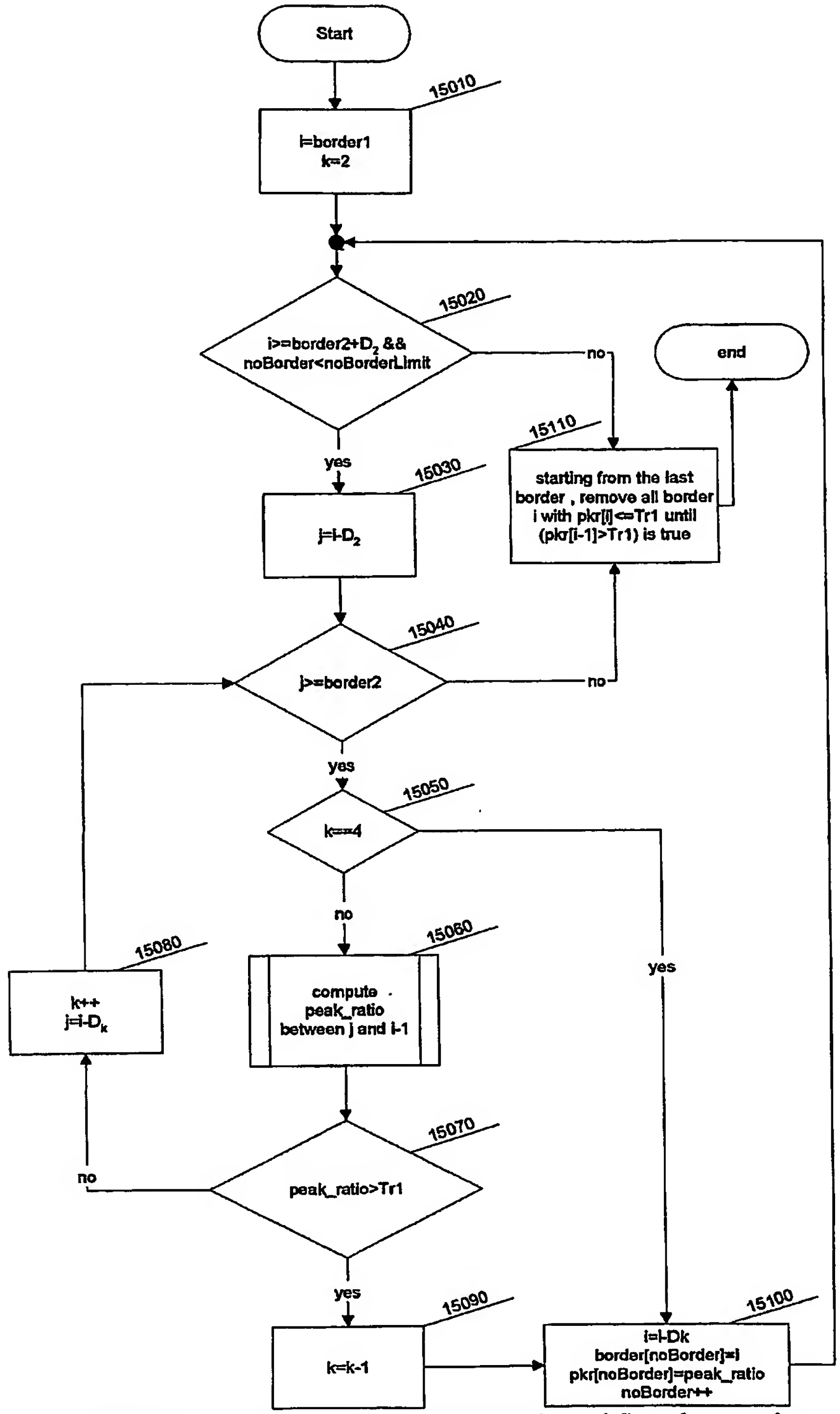

Figure 15: The flowchart of the Backward Search operation

【図１６】

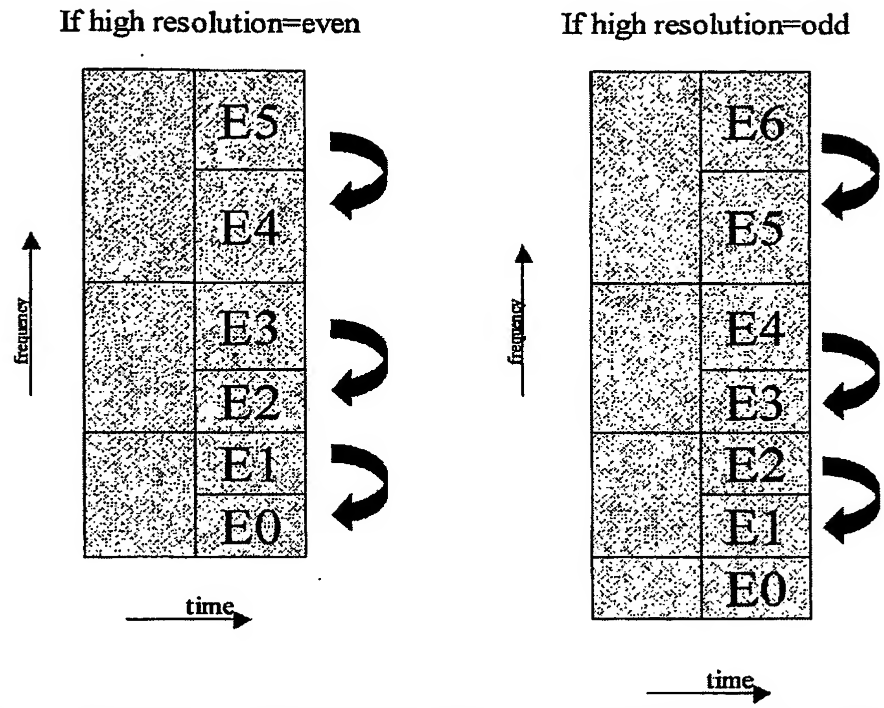

Figure 16:　Illustration for the frequency resolution determination part of the invention

【図１７】

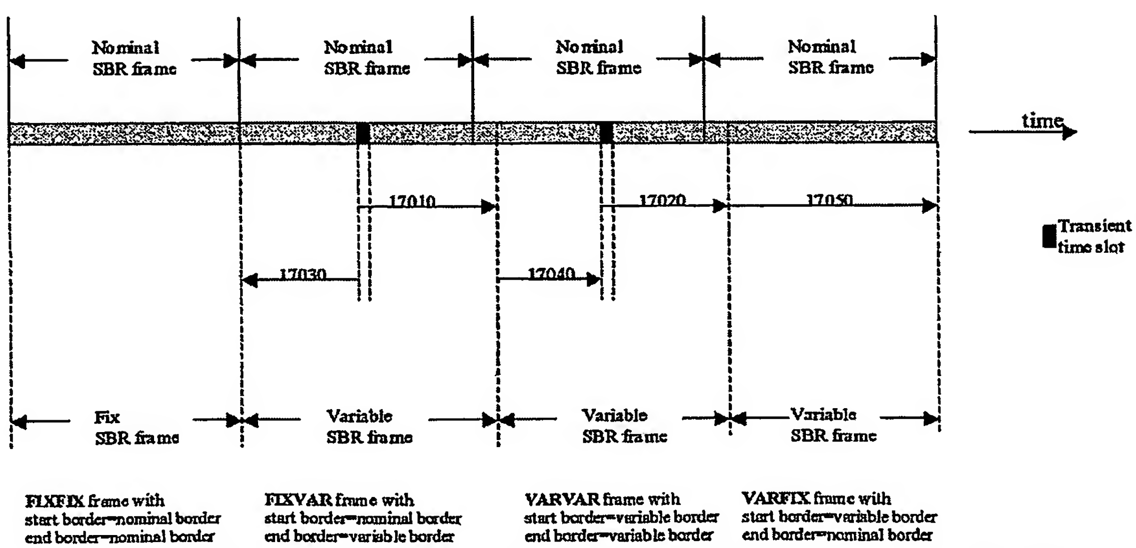

Figure 17:　Employment of the three search operations in various parts of the four frame types.

【図１８】

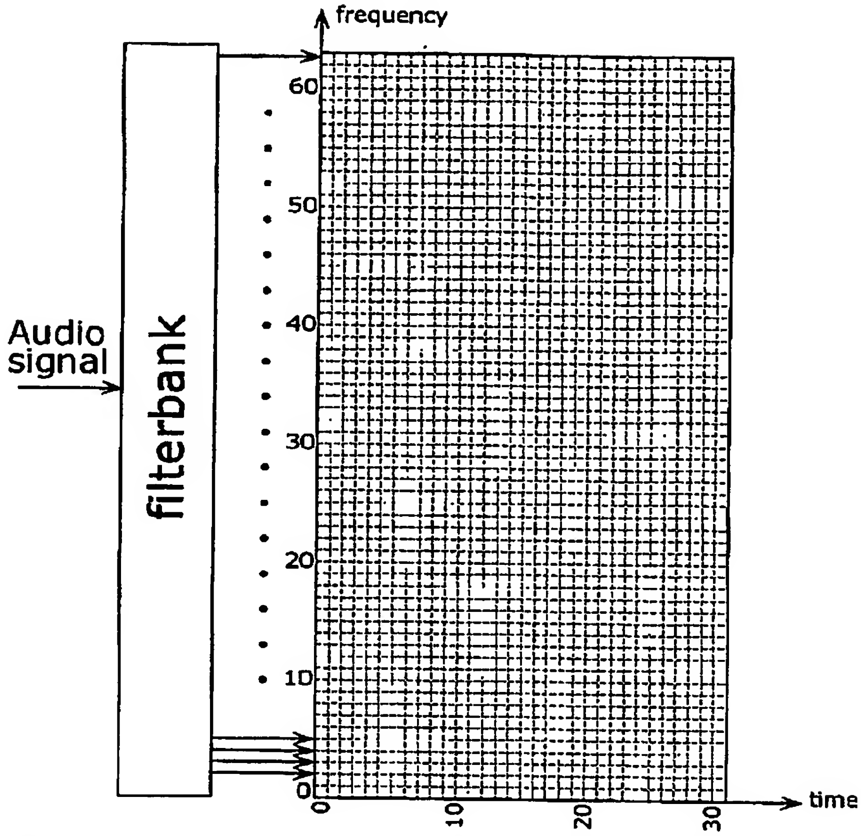

Figure 18 : A tylcal time/frequency grid representation for audio coding

【書類名】外国語要約書

5 ABSTRACT

5.1 Issue

For subband coding based bandwidth extension methods such as the SBR, proper segmentation in both the time and frequency direction is important to prevent low-energy areas from sharing the same average energy value with the large-energy areas. Otherwise, wrongful amplification might occur at the decoder, which might lead to audible artefacts.

5.2 Solution

Like in the prior art, the frame type for the current SBR frame is determined according to the type of end border of the previous frame, as well as the presence of a transient in the current SBR frame. The start border is determined according to the end border of the previous SBR frame. For a FIXFIX frame, a low time-resolution setting is used. For a FIXVAR or a VARVAR frame, a search for intermediate borders is conducted in the region between the transient and maximum allowed end border location. The end border is also determined at this stage. If there is excess capacity for more borders, another search is conducted in the region between the transient and the start border. For a VARFIX frame, only one search needs to be conducted, in the whole region flanked by a variable start border and a fixed end border. All of the above are accomplished with two Forward Search operations and one Backward Search operation. They employ the same principle, which is based on evaluating the signal variation of a time segment, but with minor variations to suit the scenarios in which they are applied. To determining the frequency resolution, a high resolution is first assumed and average energies are computed for each frequency band. For every pair of frequency bands flanked by the low-resolution borders, the ratio of energies is computed. If the minimum of all energy differences computed for the entire time segment exceeds a pre-determined threshold, a high-frequency resolution is adopted. Otherwise, a low-frequency resolution is adopted. The method applies a stricter criterion for the adoption of high frequency resolution in this region.

認定・付加情報

特許出願の番号	特願２００３－３４８６２１
受付番号	５０３０１６７１３８７
書類名	特許願
担当官	塩野　実　　　２１５１
作成日	平成１５年１２月　４日

<認定情報・付加情報>

【提出日】　平成15年10月 7日

特願２００３－３４８６２１

出　願　人　履　歴　情　報

識別番号　　　　　［０００００５８２１］

１．変更年月日　　　１９９０年　８月２８日
　　［変更理由］　　新規登録
　　　　住　所　　　大阪府門真市大字門真１００６番地
　　　　氏　名　　　松下電器産業株式会社